Data Sci. Eng. (2016) 1(2):51–53
DOI 10.1007/s41019-016-0010-4

EDITORIAL

Introduction to the Special Issue

Jianzhong Li[1]

Received: 11 June 2016 / Accepted: 14 June 2016 / Published online: 22 July 2016

With the rapid development of information technologies and data acquisition technologies, the era of big data has arrived. In the recent years, big data have been accumulated by many countries, and the size of data has exceeded PB (10^{15} bytes) in many areas such as energy, manufacturing, transportation, service industry, education and public health. The enormous values of big data can be summarized as follows.

(1) *Accelerating economic development* In 2012, Gartner, a world-famous international consulting organization, released a report forecasting that the global big data-related industrial scale will reach $232 billion in 2016. A Mckinsey Global Institute report forecasts that big data will bring an economic value of $300 billion annually only in the medical industry. In China, many companies such as Alibaba and Baidu have gained a great amount of economic benefit by querying, mining and analyzing big data.

(2) *Promoting social progress* Big data play an important role in social progress, which can be witnessed by the following facts. Financial big data can help the government on the aspects of decision-making. Big data from the areas of energy, transportation, environment and geography can help with city management, conserving energy, intelligent traffic control and environment monitoring. Big health data can be utilized to monitor the running status of medical systems and the trends of people's health, to lower healthcare costs and to improve health care standard and health conditions of the whole society. Big data from education can help with promoting the effectiveness and efficiency of education.

(3) *Enhancing social security and stability* Through querying, mining and analyzing big data over the Internet, social trends can be timely observed and analyzed, and the outbursts of significant sensitive events can be predicted, so that the government can promptly react to enhance social security and stability. Furthermore, big data can provide a large number of employment opportunities, which also benefits social stability. A new Mckinsey Global Institute report finds that 140,000–190,000 staff members and 1.5 million data managers are in demand in the big data industry merely in the USA.

(4) *Renovating scientific research methods* Big data pave a new road for scientific research, which accelerates the scientific revolution. Scientific research methods will transform from the previous hypothesis-driven ones to the data-driven ones. By querying, mining and analyzing big data, scientists can find novel phenomenon and laws. For instance, Sloan Digital Sky Survey Database has become a core resource for astronomical research, where astronomers have discovered a large number of new astronomical phenomenon and laws. It is popular nowadays to perform biology research toward biological big data and a whole set of methodology has been developed.

The aforementioned enormous values of big data are embodied in practical big data applications which bring a large number of big data-related challenging research issues. This Special Issue is to demonstrate the value of big data through practical applications and to reveal the big data-related research issues brought by the practical applications. The Special Issue consists of six papers.

The special issue begins with the paper "Medical big data: neurological diseases diagnosis through medical data analysis" by Siuly Siuly and Yanchun Zhang. This paper concerns the diagnosis of the neurological diseases based big medical data. The challenges of medical big data handing and the con-

✉ Jianzhong Li
lijzh@hit.edu.cn

[1] Harbin Institute of Technology, Harbin, China

cept of the computerized diagnosis systems are introduced in the paper. The paper also provides a survey of the developed computerized diagnosis methods using big medical data in the area of neurological diseases diagnosis.

The second paper is "Software-Defined Storage Based Data Infrastructure Supportive of Hydroclimatology Simulation Containers: A Survey" by Wonjun Lee and Sanjiv Kumar. This paper focuses on the Hydroclimatic research using big data. Hydroclimatic research requires highly intensive resources of big data and computation to perform simulations. To address the problems of performing simulations on big data in hydroclimatic research, Software-Defined Storage and Containers have been introduced. This paper introduces the hydroclimatic experiments and their fast deployment and high portability and scalability in Container. The Software-Defined Storage technologies in connection with the hydroclimatic simulations are also introduced in this paper. This paper also discusses how the Software-Defined Storage data infrastructure strengthens Containers in terms of the flexibility and scalability of data handling and storage.

The third paper is "A Data Driven Evaluation for Insider Threats" by Yuqing Sun, Haoran Xu, Elisa Bertino and Chao Sun. This paper tackles a data security problem, the insider threat problem, by auditing user behaviors. Insiders are often legal users who are authorized to access systems and data. If they misuse their privileges, it would bring great threat to system security. In this paper, a data-driven method to detect malicious insiders is proposed. First, a normal pattern of a user group and a routine behavior pattern of each user are learned for historical assessment. Then, users are evaluated against the patterns by probabilistic methods. The deviation degree is adopted as an evidence to justify an anomaly. This paper also recognizes the abnormal activities that often make a user behavior much deviation, which can help an administrator revisit security policies or update activity weights in assessment.

The fourth paper is "Landmark-based Route Recommendation with Crowd Intelligence" by Bolong Zheng, Han Su, Kai Zheng and Xiaofang Zhou. This paper is to solve the route recommendation problem by combining the mining of big trajectory data or map services and the crowds' knowledge. Route recommendation is one of the most widely used location-based services nowadays. Given a pair of user-specified origin and destination, a route recommendation service aims to provide users with the routes of the best traveling experience according to a given criteria. However, even the routes recommended by the big-thumb service providers can deviate significantly from the ones travelled by experienced drivers, which motivates previous research that leverages crowds' knowledge to improve the recommendation quality. Thus, this paper proposes a two-layer system that receives user's request from mobile client specifying the source and destination, processes the request on the server and finally returns the verified best routes to the user. The proposed system comprises two modules: traditional route recommendation (TR) and crowd-based route recommendation (CR). The TR module firstly processes user's request by trying to evaluating the quality of candidate routes obtained from big trajectory data or map services. The CR module generates a crowd-sourcing task when the TR module cannot judge the quality of candidate routes, and return the best route based on the feedbacks of human workers of the system. To reduce the latency of the route recommendation, this paper also proposes some strategies to reuse existing answers (truths) to deal with newly coming queries more efficiently.

The fifth paper is "UniClip: Leveraging Web Search for Universal Clipping of Articles on Mobile" by Ruihua Song, Kazutoshi Umemoto, Jian-Yun Nie, Xing Xie, Katsumi Tanaka and Yong Rui. This paper considers the problem of searching big Web data and addresses the difficulty of clipping articles from mobile apps. A service called UniClip that allows a user to save the full content of an article by snapping a screenshot part of it is proposed. UniClip leverages a huge amount of indexed Web data to mine the article by starting with a snapped screenshot. Three approaches are presented to solve three challenges, which are how to represent a screenshot, how to formulate effective queries for retrieving a full article and how to rank the best URL at the top from multiple search result lists.

The last paper is "A Storytelling driven Framework for Cultural Heritage Dissemination" by Stefano Valtolina. This paper investigates the problem of disseminating Cultural Heritage in digitalized museums. A new dissemination framework for Cultural Heritage is proposed. It shows possible solutions for small and medium museums to cooperate/collaborate in the creation of exhibitions, makes possible new data-based communication strategies which are able to combine contents belonging to different cultural archives and accessed through an ontology-based integration and discovery mechanism, and fosters new data sharing and distribution policies that preserve intellectual property rights. The framework redefines the concept of digital storytelling with the aim of increasing the participation of domain experts in the dissemination of Cultural Heritage. Recommender and digital right management services are also provided in the framework to authors and users for helping them in the creation, personalization and navigation of stories and for guaranteeing the adoption of suitable sharing and distribution policies.

At the end, I want to thank professors Elisa Bertino and Lizhu Zhou for their encouragement and help. I am also thankful for the authors, not only of the accepted papers, but also all of those who submitted their works. The special issue would be not possible without the enthusiasm of the authors.

Finally, thanks are due to the reviewers who worked hard to evaluate the papers and assist the authors in the process of making their papers the strong contributions that they are. I hope the readers will be well rewarded for reading the papers in this special issue.